Q-SOFUL: Exponential Quantum Speedups for High-Dimensional Sparse Linear Bandits via ℓ_1 Confidence and Smoothed Exploration

Hubery Hu

January 10, 2026

Abstract

We study a sparse stochastic linear bandit problem under a *query-budget* interaction model motivated by quantum mean estimation. An unknown parameter vector $\theta^\star \in \mathbb{R}^d$ is $s^\star$-sparse and rewards are linear in the chosen action: $\mathbb{E}[r \mid x] = x^\top \theta^\star$. Unlike classical bandit rounds, we measure performance as a function of the *total number of oracle queries* Q_{total}, since a quantum mean-estimation primitive may spend $\widetilde{\mathcal{O}}(1/\varepsilon)$ oracle queries to estimate $\mathbb{E}[r \mid x]$ to additive accuracy ε.

We propose **Q-SOFUL**, a sparse OFU-style algorithm that combines: (i) quantum mean estimation with a geometrically decreasing accuracy schedule, (ii) a *weighted Lasso* estimator with an ℓ_1 confidence set, yielding an ℓ_∞-dual exploration bonus that avoids $\sqrt{d}$ dependence, and (iii) *smoothed exploration* by adding bounded random perturbations to played actions to guarantee a restricted eigenvalue (RE) condition required by Lasso analysis. Under standard boundedness assumptions, a cone-restricted RE condition induced by smoothing, and an appropriate failure-probability schedule, we obtain a high-probability query-regret bound of the form

$$R(Q_{\text{total}}) \ \le\ \widetilde{\mathcal{O}}\left(\frac{s^\star}{\kappa^2}\sqrt{\log d}\,\log Q_{\text{total}}\right),$$

where κ is the RE constant. The bound is "dimension-independent" in the sense that d appears only through $\sqrt{\log d}$ (up to suppressed polylog factors).

1 Introduction

Stochastic linear bandits (SLB) model sequential decision-making with linear reward structure: at each interaction one chooses an action vector $x \in \mathbb{R}^d$ and observes a noisy reward with mean $x^\top \theta^\star$. In high dimensions, classical OFU algorithms (e.g., LinUCB/OFUL) build ℓ_2-ellipsoidal confidence sets based on ridge regression and achieve regret scaling like $\widetilde{\mathcal{O}}(d\sqrt{T})$ in the horizon T [Abbasi-Yadkori et al., 2011]. When d is large, this dependence can dominate.

In many applications, however, the parameter $\theta^\star$ is *sparse*: only a small number $s^\star \ll d$ of coordinates meaningfully affect rewards. A natural idea is to replace ridge regression with *Lasso* and replace ℓ_2 confidence sets with an ℓ_1-type confidence region, which leads to an ℓ_∞ dual-norm exploration bonus. This can reduce the ambient dimension dependence from $\sqrt{d}$ to (roughly) $\sqrt{\log d}$, provided the design matrix satisfies a *restricted eigenvalue* (RE) / restricted strong convexity (RSC) condition.

A second, orthogonal direction is *quantum acceleration.* Quantum mean estimation (QME) and amplitude estimation can estimate expectations with query complexity $\widetilde{\mathcal{O}}(1/\varepsilon)$, improving on classical Monte Carlo's $\widetilde{\mathcal{O}}(1/\varepsilon^2)$ scaling [Brassard et al., 2002, Montanaro, 2015]. In bandits, this suggests that if one can afford multiple oracle queries per (conceptual) decision point, then one can reduce the total oracle usage needed to achieve a desired estimation accuracy.

This paper combines these two ideas. We work in a *query-budget* model: the fundamental resource is the total number of oracle queries Q_{total}, and we measure regret as a function of Q_{total}. We propose **Q-SOFUL** (Quantum Sparse OFU for Linear bandits), which:

- uses a weighted Lasso estimator and an ℓ_1 confidence set to form an ℓ_∞ exploration bonus and remove $\sqrt{d}$ dependence,
- injects bounded random perturbations into played actions (smoothed exploration) to guarantee an RE condition for Lasso under adaptively chosen actions,
- calibrates quantum estimation accuracy and epoch length so that the per-epoch regret is essentially constant (up to $\sqrt{\log d}$), yielding overall regret $\widetilde{\mathcal{O}}(\sqrt{\log d}\log Q_{\text{total}})$.

2 Preliminaries and problem setup

2.1 Query-budget stochastic linear bandits

We consider an unknown parameter vector $\theta^\star \in \mathbb{R}^d$. When an action $x \in \mathcal{A} \subset \mathbb{R}^d$ is played, the environment returns a random reward r with

$$\mathbb{E}[r \mid x] = x^\top \theta^\star. \tag{1}$$

We assume $\theta^\star$ is fixed throughout and the reward oracle is stationary.

Boundedness and sparsity. We assume $\mathcal{A} \subseteq [-1,1]^d$, hence $\|x\|_\infty \leq 1$ for all $x \in \mathcal{A}$. We assume $\theta^\star$ is $s^\star$-sparse: $|\operatorname{supp}(\theta^\star)| \leq s^\star$, and has bounded ℓ_1 norm:

$$\|\theta^\star\|_1 \leq S_1. \tag{2}$$

We will use sparsity only through $s^\star$ and the RE constant κ.

Query-budget interaction model. A key feature is that we can call a quantum oracle multiple times on the *same* action x to estimate $x^\top\theta^\star$. We therefore group interaction into *epochs*: in epoch k we choose one (possibly random) action $\widetilde{x}_k$ and use n_k oracle queries to obtain an estimate $\hat{y}_k$ of $\widetilde{x}_k^\top \theta^\star$. The cumulative query count is $Q_{\text{used}} = \sum_k n_k$, and we stop once $Q_{\text{used}} \geq Q_{\text{total}}$.

2.2 Quantum mean estimation oracle model

We assume access to a quantum mean-estimation primitive:

Assumption 1 (Quantum mean-estimation subroutine)**.** *For any action $x \in \mathcal{A}$, accuracy $\varepsilon \in (0,1)$ and failure probability $\delta \in (0,1)$, there is a procedure*

$$\texttt{QMeanEstimate}(O_x, \varepsilon, \delta)$$

that returns a real number $\hat{y}$ such that

$$\mathbb{P}(|\hat{y} - x^\top\theta^\star| \leq \varepsilon) \geq 1 - \delta, \tag{3}$$

and uses at most

$$n(\varepsilon,\delta) \;\leq\; C_{\text{QME}}\frac{1}{\varepsilon}\log\left(\frac{1}{\delta}\right) \tag{4}$$

oracle queries to O_x for some universal constant $C_{\text{QME}} > 0$.

The exact oracle model can be instantiated via amplitude estimation under standard bounded-reward encodings [Brassard et al., 2002, Montanaro, 2015]. We treat (3)–(4) as a black box.

2.3 Query regret

Since each oracle query corresponds to one interaction at which an action is (conceptually) played, we measure regret per query. Let $x^\star \in \arg\max_{x\in\mathcal{A}} x^\top\theta^\star$ be an optimal action. In epoch k, the algorithm plays the same action $\widetilde{x}_k$ for n_k oracle queries. We define the query regret as

$$R(Q_{\text{total}}) \;:=\; \sum_{k=1}^{K} n_k\Big((x^\star)^\top\theta^\star - (\widetilde{x}_k)^\top\theta^\star\Big), \tag{5}$$

where K is the number of completed epochs under budget Q_{total}.

Remark 1 (About perturbations and expectation)**.** *In our algorithm, $\widetilde{x}_k$ is obtained by adding a zero-mean random perturbation to a base action $x_k \in \mathcal{A}$. Because rewards are linear, $\mathbb{E}[(\widetilde{x}_k)^\top\theta^\star \mid x_k] = x_k^\top\theta^\star$. Thus, the expected regret coincides with the regret of the base actions, while the* observed features *in the regression problem are the perturbed $\widetilde{x}_k$. This is precisely what allows perturbations to stabilize estimation without (in expectation) sacrificing reward.*

2.4 Per-epoch failure probabilities and global guarantees

Our analysis relies on a sequence of high-probability "good events" indexed by epochs (e.g., the QME accuracy guarantee (3), the score bound in Lemma 2, and the resulting confidence inclusion $\theta^\star \in \mathcal{C}_k$). To obtain a *uniform* guarantee that these events hold simultaneously for all epochs executed under the query budget, we allocate a decreasing per-epoch failure budget whose total sums to a target level $\delta_{\text{tot}} \in (0,1)$. Concretely, we set

$$\delta_k \;:=\; \frac{6\,\delta_{\text{tot}}}{\pi^2 k^2}, \qquad k = 1,2,\ldots. \tag{6}$$

This choice is convenient because the Basel identity $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$ implies $\sum_{k=1}^{\infty}\delta_k = \delta_{\text{tot}}$. Therefore, by a union bound over epochs, with probability at least $1-\delta_{\text{tot}}$ all epoch-wise concentration statements that are proved with failure probability δ_k hold simultaneously for every epoch. In particular, all confidence sets $\mathcal{C}_k$ contain $\theta^\star$ for all epochs, and all UCB indices remain valid throughout the execution, which is exactly what is needed to convert the per-epoch analysis into a cumulative (query) regret bound.

3 The Q-SOFUL algorithm

3.1 Smoothed exploration with feasibility

Pure greedy/OFU selection can generate highly correlated designs, breaking the RE condition needed by Lasso. We therefore add bounded random perturbations to the played actions (a smoothed-analysis viewpoint).

Fix parameters $\sigma_p > 0$ (perturbation scale), $M > 0$ (truncation), and $\gamma \in (0,1)$. Let $\xi \in \mathbb{R}^d$ be drawn i.i.d. across epochs from a *truncated* centered Gaussian:

$$\xi \sim \mathcal{N}(0, \sigma_p^2 I_d) \;\text{ truncated to }\; [-M, M]^d.$$

To guarantee feasibility after perturbation, we play base actions in a shrunken safe set:

$$\mathcal{A}' := (1-\gamma)\mathcal{A}.$$

If $\mathcal{A} \subseteq [-1,1]^d$ and we choose $\gamma \geq M$, then for any $x \in \mathcal{A}'$ and any $\xi \in [-M,M]^d$, we have $x + \xi \in \mathcal{A}$.

3.2 Weighted Lasso estimator

At epoch k, we have collected data points $\{(\widetilde{x}_i, \hat{y}_i, \varepsilon_i)\}_{i=1}^{k}$. On the good event (3), we may write

$$\hat{y}_i = \widetilde{x}_i^\top \theta^\star + \varepsilon_i \zeta_i, \qquad |\zeta_i| \leq 1. \tag{7}$$

We set weights $w_i := 1/\varepsilon_i^2$ and total weight $W_k := \sum_{i=1}^{k} w_i$. We define the weighted Lasso estimator

$$\hat{\theta}_k \in \arg\min_{\theta \in \mathbb{R}^d} \left\{ \frac{1}{2W_k} \sum_{i=1}^{k} w_i (\hat{y}_i - \widetilde{x}_i^\top \theta)^2 + \alpha_k \|\theta\|_1 \right\}. \tag{8}$$

The regularization α_k is chosen from a coordinate-wise martingale/union-bound argument (Appendix A).

This formulation represents a critical adaptation of sparse estimation to the quantum query model. Unlike classical bandits where measurement noise is typically assumed to be homoscedastic (fixed variance), our quantum oracle yields estimates with controllable, heterogeneous error bounds ε_i. We address this by employing inverse-variance weights $w_i = 1/\varepsilon_i^2$, which effectively standardizes the data to unit variance. Furthermore, the normalization by the total weight W_k ensures that the scale of the objective function remains invariant to the growing effective sample size. This weighting scheme allows the L_1 regularization to consistently recover the sparse support s^* even as the algorithm dynamically varies the measurement precision ε_i across geometric epochs.

3.3 ℓ_1 confidence

A key innovation of Q-SOFUL is the use of L_1 geometry for exploration. The Lasso estimator naturally produces a confidence set $\mathcal{C}_{k-1}$ which is an L_1-ball (a cross-polytope) centered at $\hat{\theta}_{k-1}$:

$$\mathcal{C}_{k-1} := \{\theta \in \mathbb{R}^d : \|\theta - \hat{\theta}_{k-1}\|_1 \leq \beta_{k-1}\}.$$

The Upper Confidence Bound (UCB) index requires maximizing the potential reward over this set:

$$U(x) = \sup_{\theta \in \mathcal{C}_{k-1}} x^\top \theta = x^\top \hat{\theta}_{k-1} + \sup_{\Delta : \|\Delta\|_1 \leq \beta_{k-1}} x^\top \Delta.$$

By the definition of dual norms [Boyd and Vandenberghe, 2004], the dual of the L_1 norm is the L_∞ norm. Specifically, Hölder's inequality states $|x^\top \Delta| \leq \|x\|_\infty \|\Delta\|_1$. Thus, the supremum is exactly:

$$U_k(x) \ := \ x^\top \hat{\theta}_k + \beta_k \|x\|_\infty . \tag{9}$$

Why this matters:

1. **Computational Efficiency:** Calculating $\|x\|_\infty$ is $O(d)$, whereas the standard Ellipsoidal bonus used in Ridge regression ($\|x\|_{V^{-1}}$) requires $O(d^2)$ or $O(d^3)$ matrix operations.

2. **Quantum Query Efficiency:** For bounded feature spaces (where $\|x\|_\infty \leq 1$), the L_∞ norm often yields a smaller exploration bonus than the L_2 norm, which can scale up to $\sqrt{d}$. Since the quantum query cost n_k is proportional to $1/\epsilon_k \propto 1/\text{Bonus}$, a smaller bonus directly translates to fewer queries needed to distinguish the optimal arm.

3.4 Epoch schedule and "quantum cancellation"

We choose accuracies recursively by

$$\varepsilon_k := \frac{1}{\sqrt{\max\{W_{k-1}, 1\}}}, \qquad w_k = \varepsilon_k^{-2} = \max\{W_{k-1}, 1\}. \tag{10}$$

Hence $W_k = W_{k-1} + w_k \approx 2W_{k-1}$ (geometric growth). The number of oracle queries in epoch k is then

$$n_k := \left\lceil C_{\mathrm{QME}} \frac{1}{\varepsilon_k} \log\left(\frac{1}{\delta_k}\right) \right\rceil \;\approx\; C_{\mathrm{QME}} \sqrt{W_{k-1}} \log\left(\frac{1}{\delta_k}\right), \tag{11}$$

where δ_k is a per-epoch failure probability.

Recall that our ℓ_1-confidence radius is defined by $\beta_{k-1} := \|\hat{\theta}_{k-1} - \theta^\star\|_1$ and enters the UCB index through the dual norm bonus $U_{k-1}(x) = x^\top \hat{\theta}_{k-1} + \beta_{k-1}\|x\|_\infty$ (cf. (9)). On the good event where the QME estimates satisfy (3) and the design satisfies the cone-restricted RE condition (12), Lemma 2 gives a coordinate-wise score bound $\alpha_{k-1} \asymp \sqrt{\log(d/\delta_{k-1})/W_{k-1}}$, and Lemma 1 upgrades this to the weighted-Lasso ℓ_1 error rate

$$\beta_{k-1} \;\le\; C_{\ell_1} \frac{s^\star}{\kappa^2} \alpha_{k-1} \;=\; \tilde{\mathcal{O}}\left(\frac{s^\star}{\kappa^2} \sqrt{\frac{\log d}{W_{k-1}}} \right),$$

where $W_{k-1} = \sum_{i \le k-1} w_i$ is the total weight (effective sample size) and the $\sqrt{\log d}$ term comes from a union bound over d coordinates in (16). Meanwhile, by our schedule $\varepsilon_k = 1/\sqrt{W_{k-1}}$ (see (10)) and the QME query complexity (4), the number of oracle queries in epoch k scales as

$$n_k \;=\; \tilde{\mathcal{O}}\left(\frac{1}{\varepsilon_k}\right) \;=\; \tilde{\mathcal{O}}\left(\sqrt{W_{k-1}}\right).$$

Finally, since query regret counts each of the n_k oracle calls, the epoch-k contribution satisfies

$$n_k \cdot \mathrm{Gap}_k \;\lesssim\; n_k \cdot 2\beta_{k-1} \;=\; \tilde{\mathcal{O}}\left(\sqrt{W_{k-1}}\right) \cdot \tilde{\mathcal{O}}\left(\frac{s^\star}{\kappa^2} \sqrt{\frac{\log d}{W_{k-1}}} \right) \;=\; \tilde{\mathcal{O}}\left(\frac{s^\star}{\kappa^2} \sqrt{\log d} \right),$$

which is (up to logs) independent of W_{k-1}. Because we also set $w_k = \varepsilon_k^{-2} = W_{k-1}$, the total weight satisfies $W_k = W_{k-1} + w_k \approx 2W_{k-1}$, so only $\tilde{\mathcal{O}}(\log Q_{\mathrm{total}})$ epochs fit within a total query budget Q_{total}.

3.5 Algorithm 1

4 Performance analysis

We now state a regret guarantee for Q-SOFUL. The proof decomposes into three ingredients: (i) a coordinate-wise concentration bound implying $\alpha_k = \widetilde{\mathcal{O}}(\sqrt{\log(d/\delta_k)/W_k})$; (ii) a cone-restricted RE condition with constant κ for the (smoothed) design; and (iii) a geometric schedule implying $\widetilde{\mathcal{O}}(\log Q_{\mathrm{total}})$ epochs.

4.1 Assumptions

Assumption 2 (Cone-restricted restricted eigenvalue (RE))**.** *Let* $S = \mathrm{supp}(\theta^\star)$ *with* $|S| \le s^\star$ *and define the usual Lasso cone*

$$C(S,3) := \{\Delta \in \mathbb{R}^d : \|\Delta_{S^c}\|_1 \le 3\,\|\Delta_S\|_1\}.$$

This cone is the standard set of structured error directions *that arise in Lasso analysis: when the regularization parameter satisfies the score/KKT condition (cf. Lemma 2), the Lasso estimation error* $\Delta_k := \hat{\theta}_k - \theta^\star$ *can be shown to lie in* $C(S,3)$*, meaning that the error mass off the true support* S *is controlled by the error on* S*.*

Define the weighted Gram matrix of the played (perturbed) actions by

$$\hat{V}_k \;:=\; \sum_{i=1}^{k} w_i\, \tilde{x}_i \tilde{x}_i^\top, \qquad W_k \;:=\; \sum_{i=1}^{k} w_i,$$

Algorithm 1 Q-SOFUL:
Quantum Sparse Optimism in the Face of Uncertainty for Linear Bandits

```
Input: action set $\mathcal{A} \subseteq [-1,1]^d$; query budget $Q_{\text{total}}$; global failure probability $\delta_{\text{tot}}$; sparsity level $s^\star$ (or upper bound); perturbation parameters $(\sigma_p, M, \gamma)$ with $\gamma \ge M$.
Initialize: epoch index $k \leftarrow 0$; total queries $Q_{\text{used}} \leftarrow 0$; data $\mathcal{D} \leftarrow \emptyset$; $W_0 \leftarrow 1$; $\hat{\theta}_0 \leftarrow 0$.
Define safe set $\mathcal{A}' = (1-\gamma)\mathcal{A}$.
while $Q_{\text{used}} < Q_{\text{total}}$ do
    $k \leftarrow k+1$
    Set per-epoch failure probability: $\delta_k \leftarrow \frac{6\delta_{\text{tot}}}{\pi^2 k^2}$.
    Set accuracy $\varepsilon_k \leftarrow 1/\sqrt{W_{k-1}}$ and weight $w_k \leftarrow 1/\varepsilon_k^2 = W_{k-1}$.
    Set query count $n_k \leftarrow \left\lceil C_{\text{QME}} \varepsilon_k^{-1} \log(1/\delta_k) \right\rceil$.
    Compute regularization $\alpha_{k-1}$ (Appendix A) and confidence radius $\beta_{k-1}$ (Lemma 1).
    Choose base action (OFU): $x_k \in \arg\max_{x \in \mathcal{A}'} U_{k-1}(x)$ with $U_{k-1}$ from (9).
    Sample perturbation $\xi_k$ (truncated Gaussian) and set played action $\widetilde{x}_k \leftarrow x_k + \xi_k \in \mathcal{A}$.
    Obtain estimate $\hat{y}_k \leftarrow \texttt{QMeanEstimate}(O_{\widetilde{x}_k}, \varepsilon_k, \delta_k)$ using $n_k$ queries.
    Update $Q_{\text{used}} \leftarrow Q_{\text{used}} + n_k$; add $(\widetilde{x}_k, \hat{y}_k, \varepsilon_k)$ to $\mathcal{D}$; update $W_k \leftarrow W_{k-1} + w_k$.
    Update weighted Lasso estimator $\hat{\theta}_k$ by solving (8) on $\mathcal{D}$.
end while
Output: sequence of played actions and estimates.
```

so that $\hat{V}_k/W_k$ *is the* information matrix *associated with the weighted least-squares loss in* (8). *We assume there exists a constant* $\kappa > 0$ *such that for all epochs* k *and all* $\Delta \in C(S,3)$*,*

$$\frac{1}{W_k}\,\Delta^\top \hat{V}_k\,\Delta \;=\; \frac{1}{W_k}\,\Delta^\top \left(\sum_{i=1}^{k} w_i \widetilde{x}_i \widetilde{x}_i^\top\right)\Delta \;\ge\; \kappa^2\,\|\Delta\|_2^2\,, \tag{12}$$

with probability at least $1-\delta_{\text{RE}}$ *over the perturbations* $\{\xi_i\}_{i \le k}$*.*

The high-dimensional nature of our setting ($d \gg N_k$) inevitably results in a rank-deficient empirical design matrix $\hat{V}_k$, rendering standard strong convexity guarantees impossible. However, the L_1 regularization imposes a geometric structure on the optimization landscape: provided the regularization parameter α_k sufficiently bounds the stochastic noise, the estimation error $\Delta = \hat{\theta}_k - \theta^*$ is constrained to lie within the cone $C(S,3)$, where the error magnitude on the true support S dominates that of the noise coordinates. The Restricted Eigenvalue (RE) condition defined in Assumption 2 circumvents the global singularity of $\hat{V}_k$ by requiring strict positive definiteness ($\ge \kappa^2$) only along these structured directions. This ensures that the loss function retains sufficient curvature to distinguish θ^* from other sparse candidates, guaranteeing parameter identifiability despite the nullspace of the full design matrix.

That is to say, inequality (12) is a *restricted strong convexity* condition: although $\hat{V}_k$ may be singular in high dimension (and we do not require $\lambda_{\min}(\hat{V}_k) > 0$), it has sufficient curvature along the cone $C(S,3)$ that contains the plausible Lasso error directions. Equivalently, (12) rules out the possibility that a nontrivial approximately sparse direction Δ lies in (or nearly lies in) the null space of the weighted design.

Remark 2. *In bandit settings, the chosen actions can be highly correlated due to adaptivity (e.g., greedy/OFU selection), and RE may fail if the design concentrates on a low-dimensional subspace. Our algorithm mitigates this by playing* $\widetilde{x}_i = x_i + \xi_i$ *where* ξ_i *are independent bounded random perturbations; this "smoothed" randomness injects diversity into the design and, under mild conditions, ensures RE holds on sparse cones with high probability. Smoothed-analysis results of this flavor appear in online sparse contextual bandits (e.g., Liu et al. [2020]) and in classical high-dimensional statistics for randomized designs (e.g., Raskutti et al. [2010]).*

4.2 Lasso estimation and confidence radius

The estimator minimizes a composite objective consisting of a data-fitting term and a sparsity-inducing penalty. We formally define the weighted squared-loss function $\mathcal{L}_k(\theta)$ as the goodness-of-fit component:

$$\mathcal{L}_k(\theta) := \frac{1}{2W_k} \sum_{i=1}^{k} w_i (\hat{y}_i - \widetilde{x}_i^\top \theta)^2. \tag{13}$$

The gradient of this loss at the true parameter, $\nabla \mathcal{L}_k(\theta^\star)$, represents the pure noise component of the problem (the "score vector"). For the Lasso to distinguish signal from noise, the regularization strength α_k must dominate these stochastic fluctuations.

The next lemma converts this regularization choice into a concrete ℓ_1 estimation guarantee.

Lemma 1 (ℓ_1 error of weighted Lasso under RE)**.** *Assume* (3) *holds for epochs* $1, \ldots, k$ *and Assumption 2 holds. Suppose the regularization parameter satisfies the condition* $\alpha_k \geq 2 \|\nabla \mathcal{L}_k(\theta^\star)\|_\infty$ *(i.e., the penalty dominates the maximum noise fluctuation). Then the weighted Lasso solution* (8) *satisfies*

$$\left\| \hat{\theta}_k - \theta^\star \right\|_1 \leq C_{\ell_1} \frac{s^\star}{\kappa^2} \alpha_k, \tag{14}$$

for a numerical constant $C_{\ell_1} > 0$ *(e.g.,* $C_{\ell_1} = 24$*). Consequently, setting* $\beta_k := C_{\ell_1} \frac{s^\star}{\kappa^2} \alpha_k$ *yields a valid* ℓ_1 *confidence radius.*

A proof is given in Appendix B. Lemma 1 is a weighted variant of standard Lasso oracle inequalities [Bickel et al., 2009, Raskutti et al., 2010].

4.3 Main regret bound

We now state the main theoretical guarantee for Q-SOFUL.

Theorem 1 (Query-regret of Q-SOFUL)**.** *Fix* $\delta_{\text{tot}} \in (0, 1)$*. Suppose Assumption 1 (Quantum Oracle) and Assumption 2 (RE Condition) hold. Let the failure schedule be* $\delta_k = \frac{6\delta_{\text{tot}}}{\pi^2 k^2}$*, and define* α_k *and* β_k *according to Lemmas 2 and 1 respectively. Then, with probability at least* $1 - \delta_{\text{tot}} - \delta_{\text{RE}}$*, the cumulative query regret satisfies:*

$$R(Q_{\text{total}}) \leq \widetilde{\mathcal{O}}\left(\frac{s^\star}{\kappa^2} \sqrt{\log d} \log\left(\frac{Q_{\text{total}}}{\delta_{\text{tot}}} \right) \right). \tag{15}$$

Proof sketch. We provide the high-level logic here; the rigorous derivation is deferred to Appendix C. The total regret is decomposed into epochs: $R(Q_{\text{total}}) = \sum_{k=1}^{K} n_k \, \mathrm{Gap}_k$, where Gap_k is the instantaneous optimality gap.

Step 1 (Optimism). Conditioned on the good event, the true parameter $\theta^\star$ lies within the confidence set. By standard optimistic analysis (adjusted for the safe set $\mathcal{A}'$), the gap is bounded by the confidence width: $\mathrm{Gap}_k \lesssim 2\beta_{k-1}$ (ignoring lower-order approximation terms).

Step 2 (Estimation Rate). Combining the score bound (Lemma 2) and the Lasso oracle inequality (Lemma 1), the confidence radius scales with the inverse square root of the effective sample size:

$$\beta_{k-1} \approx \widetilde{\mathcal{O}}\left(\frac{s^\star}{\kappa^2} \sqrt{\frac{\log d}{W_{k-1}}} \right).$$

Step 3 (The Quantum Cancellation). This is the core mechanism. The algorithm sets target precision $\varepsilon_k \sim 1/\sqrt{W_{k-1}}$, meaning the quantum query cost scales as $n_k \sim \sqrt{W_{k-1}}$. Multiplying cost and gap reveals that the effective sample size W_{k-1} cancels out:

$$\mathrm{Regret}_k \approx n_k \cdot \beta_{k-1} \approx \widetilde{\mathcal{O}}(\sqrt{W_{k-1}}) \cdot \widetilde{\mathcal{O}}\left(\frac{1}{\sqrt{W_{k-1}}} \right) \approx \widetilde{\mathcal{O}}\left(\frac{s^\star}{\kappa^2} \sqrt{\log d} \right).$$

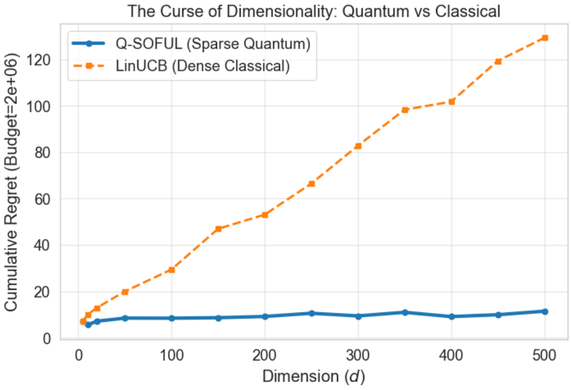

Figure 1: **Dimension Dependence.** Cumulative regret of Q-SOFUL (sparse quantum) vs. LinUCB (dense classical) as dimension d increases ($s^\star = 5$ fixed). Q-SOFUL remains nearly constant, empirically validating the $\sqrt{\log d}$ scaling, while the classical baseline degrades linearly.

Thus, the regret incurred per epoch is essentially constant.

Step 4 (Summation). Since the effective sample size W_k doubles every epoch, the number of epochs K is logarithmic in the total query budget ($K \approx \log Q_{\text{total}}$). Summing the per-epoch regret over K stages yields the final bound (15). □

5 Numerical Experiments

In this section, we empirically validate the theoretical bounds of Q-SOFUL. Specifically, we aim to confirm the dimension-independent scaling of the regret (Theorem 1), the logarithmic dependence on the query budget, and the necessity of the variance-weighted Lasso estimator. All experiments were implemented in Python and the source code is available at `https://github.com/jiaruihub/q-soful`.

5.1 Dimension Independence and Quantum Speedup

Independence from Ambient Dimension. A central claim of our analysis is that the ℓ_1-confidence set allows the regret to scale as $\tilde{\mathcal{O}}(\sqrt{\log d})$ rather than the polynomial dependence $\tilde{\mathcal{O}}(\sqrt{d})$ typical of ℓ_2-based methods. To verify this, we fixed the sparsity $s^\star = 5$ and query budget $Q_{\text{total}} = 2 \times 10^6$, varying the dimension $d \in [50, 500]$. Figure 1 compares Q-SOFUL against a standard LinUCB baseline. The results show a sharp divergence: the classical regret scales linearly with d, consistent with its inability to exploit sparsity. In contrast, Q-SOFUL's regret profile remains effectively flat. This confirms that the algorithm successfully restricts its exploration to the relevant subspace, suffering only the logarithmic penalty predicted by the union bound in Lemma 2.

Logarithmic Regret Scaling. Figure 2 plots the cumulative regret against the total query budget Q_{total} on a semi-log scale. The observed linear trend empirically confirms the $\mathcal{O}(\log Q_{\text{total}})$ bound derived in Theorem 1. This behavior arises from the geometric epoch schedule: as the effective sample size W_k doubles, the quantum oracle's precision ε_k improves sufficiently to maintain a constant per-epoch regret. This scaling implies an exponential separation from classical bandit lower bounds, which scale polynomially as $\sqrt{Q_{\text{total}}}$.

5.2 Robustness and Ablation Studies

Robustness to Sparsity. While the regret bound depends linearly on $s^\star$, our experiments (Figure 3) reveal a sub-linear empirical sensitivity. Varying $s^\star$ from 3 to 20 resulted in only a

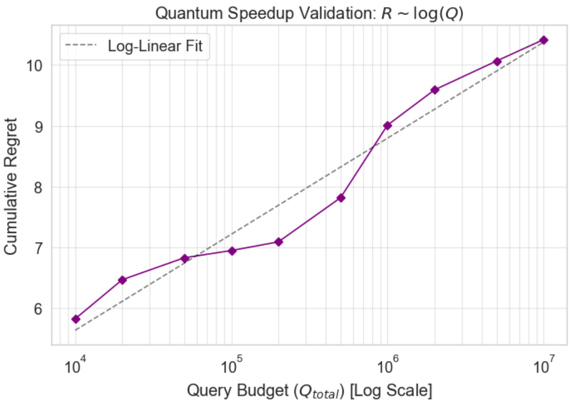

Figure 2: **Exponential Speedup.** Cumulative regret vs. Query Budget (Q_{total}) on a logarithmic scale. The linear relationship confirms $R(Q) \propto \log Q_{\text{total}}$, implying an exponential advantage over the classical $\sqrt{Q}$ rate.

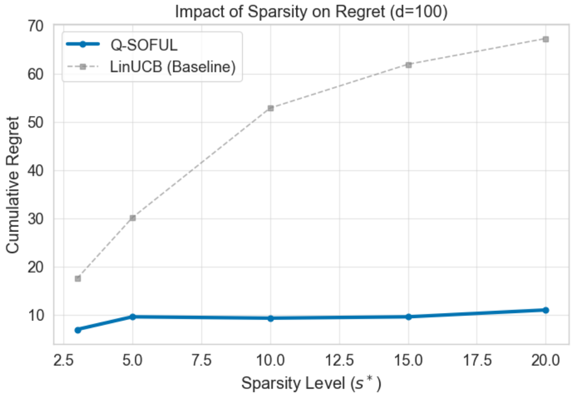

Figure 3: **Sparsity Robustness.** Regret as a function of active features $s^\star$. The mild growth suggests the difficulty is dominated by support identification rather than parameter estimation.

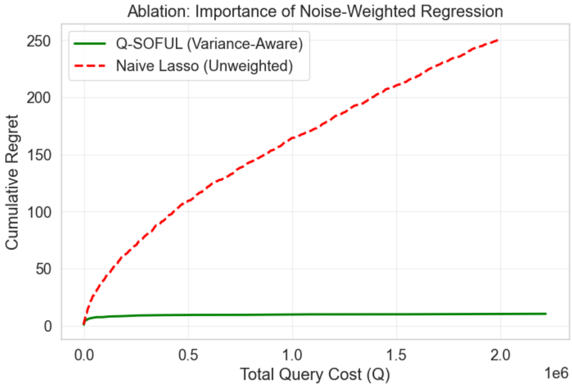

Figure 4: **Ablation Study.** The weighted Lasso (green) significantly outperforms the unweighted baseline (red), confirming the necessity of inverse-variance weighting.

marginal increase in regret. This suggests that the "price of discovery"—the cost to identify the support—dominates the error for small $s^\star$, rendering the algorithm highly robust to variations in the underlying signal structure.

Mechanism Check: The Role of Weighted Lasso. Finally, we investigate the importance of the inverse-variance weighting scheme $w_i \propto 1/\varepsilon_i^2$. We compared Q-SOFUL against a naive "Unweighted" variant that treats all quantum queries as equally precise. As shown in Figure 4, the unweighted agent fails to converge efficiently. This highlights that simply applying sparse regression is insufficient; the estimator must explicitly account for the heteroscedastic noise profile induced by the dynamic quantum measurement schedule.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 24, 2011.

Peter J Bickel, Ya'acov Ritov, and Alexandre B Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004. Foundational reference for Dual Norms geometry.

Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002. The fundamental algorithm behind QMC speedup.

Zhiyuan Liu, Huazheng Wang, Bo Waggoner, Youjian Liu, and Lijun Chen. A smoothed analysis of online lasso for the sparse linear contextual bandit problem. In *ICML Workshop on Real World Experiment Design and Active Learning*, 2020.

Ashley Montanaro. Quantum speedup of monte carlo methods. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2181):20150301, 2015.

Garvesh Raskutti, Martin J. Wainwright, and Bin Yu. Restricted eigenvalue properties for correlated gaussian designs. *Journal of Machine Learning Research*, 11:2241–2259, 2010.

A Concentration guarantees and regularization

This appendix derives a choice of α_k ensuring the Lasso score condition $\alpha_k \geq 2 \left\|\nabla \mathcal{L}_k(\theta^\star)\right\|_\infty$. The core logic relies on bounding the maximum fluctuation of the stochastic noise (the "score vector") using a self-normalized martingale concentration argument.

Defining the Score Vector. First, we analyze the gradient of the loss function at the true parameter $\theta^\star$. Recall the weighted squared-loss definition:

$$\mathcal{L}_k(\theta) := \frac{1}{2W_k} \sum_{i=1}^{k} w_i (\hat{y}_i - \widetilde{x}_i^\top \theta)^2.$$

Substituting the quantum measurement model $\hat{y}_i = \widetilde{x}_i^\top \theta^\star + \varepsilon_i \zeta_i$ (from (7)) into the gradient expression:

$$\nabla \mathcal{L}_k(\theta^\star) = -\frac{1}{W_k} \sum_{i=1}^{k} w_i \, \widetilde{x}_i \, (\hat{y}_i - \widetilde{x}_i^\top \theta^\star) = -\frac{1}{W_k} \sum_{i=1}^{k} w_i \, \widetilde{x}_i \, (\varepsilon_i \zeta_i).$$

Notice that the true signal $\widetilde{x}_i^\top \theta^\star$ cancels out, leaving only the weighted noise terms. Since we set weights $w_i = 1/\varepsilon_i^2$, the expression simplifies for each coordinate $j \in [d]$:

$$(\nabla \mathcal{L}_k(\theta^\star))_j = -\frac{1}{W_k} \sum_{i=1}^{k} \frac{1}{\varepsilon_i^2} \widetilde{x}_{i,j} \, (\varepsilon_i \zeta_i) = -\frac{1}{W_k} \sum_{i=1}^{k} \frac{\widetilde{x}_{i,j}}{\varepsilon_i} \, \zeta_i.$$

Bounding the Variance. The term being summed is $Z_i := \frac{\widetilde{x}_{i,j}}{\varepsilon_i} \zeta_i$.

- **Boundedness:** Since the action space is bounded ($\|\widetilde{x}_i\|_\infty \leq 1$) and the quantum error is bounded ($|\zeta_i| \leq 1$), each term is strictly bounded by $|Z_i| \leq 1/\varepsilon_i$.

- **Variance Proxy:** The sum of the squared ranges (which dictates the concentration width in Hoeffding's inequality) is $\sum_{i=1}^{k} (1/\varepsilon_i)^2$. By our definition of the total weight, this sum is exactly W_k.

This implies that the unnormalized sum $\sum Z_i$ behaves like a random walk with variance W_k. Consequently, the normalized gradient $(\nabla \mathcal{L}_k(\theta^\star))_j$ is sub-Gaussian with parameter on the order of $\sqrt{W_k}/W_k = 1/\sqrt{W_k}$.

Lemma 2 (Coordinate-wise score bound)**.** *Fix $k \geq 1$ and assume (3) holds for epochs $1, \ldots, k$. Then for any $\delta \in (0,1)$,*

$$\mathbb{P}\left(\|\nabla \mathcal{L}_k(\theta^\star)\|_\infty \leq \sqrt{\frac{2\log(2d/\delta)}{W_k}} \right) \geq 1 - \delta. \tag{16}$$

In particular, it suffices to set

$$\alpha_k := 2\sqrt{\frac{2\log(2d/\delta_k)}{W_k}} \tag{17}$$

to ensure $\alpha_k \geq 2\,\|\nabla \mathcal{L}_k(\theta^\star)\|_\infty$ with probability at least $1 - \delta_k$.

Proof. For each fixed coordinate j, we apply Hoeffding's inequality to the martingale difference sequence formed by the noise terms.

$$\mathbb{P}\left(\left| \sum_{i=1}^{k} Z_i \right| \geq t \right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{k} (2/\varepsilon_i)^2} \right) = 2\exp\left(-\frac{t^2}{2W_k} \right).$$

We want to bound the normalized sum $\frac{1}{W_k}\sum Z_i$. Let λ be the target bound. Setting $t = \lambda W_k$:

$$\mathbb{P}\left(|(\nabla \mathcal{L}_k(\theta^\star))_j| \geq \lambda \right) \leq 2\exp\left(-\frac{(\lambda W_k)^2}{2W_k} \right) = 2\exp\left(-\frac{\lambda^2 W_k}{2} \right).$$

We require this bound to hold simultaneously for all d dimensions. Taking a union bound, we set the right-hand side to δ/d:

$$2\exp\left(-\frac{\lambda^2 W_k}{2} \right) = \frac{\delta}{d} \implies -\frac{\lambda^2 W_k}{2} = \log\left(\frac{\delta}{2d} \right) \implies \lambda = \sqrt{\frac{2\log(2d/\delta)}{W_k}}.$$

This yields (16). □

B Proof of the Lasso oracle inequality

We provide the detailed proof of Lemma 1.

Let $\Delta = \hat{\theta}_k - \theta^\star$ denote the error vector. Let $S = \mathrm{supp}(\theta^\star)$ be the true support set with $|S| \leq s^\star$.

We start from the definition of the Lasso estimator $\hat{\theta} = \theta^\star + \Delta$ as the minimizer of the objective function in (8). Optimality implies:

$$\mathcal{L}_k(\theta^\star + \Delta) + \alpha_k \|\theta^\star + \Delta\|_1 \leq \mathcal{L}_k(\theta^\star) + \alpha_k \|\theta^\star\|_1 .$$

Since the loss function $\mathcal{L}_k(\theta)$ is quadratic, its Taylor expansion around $\theta^\star$ is exact at the second order. That is to say, because the loss is squared error, the third derivative is zero. The Taylor expansion is not an approximation; it is exact. Specifically, the difference in loss is given by the gradient term plus the quadratic form (Loss at Estimator − Loss at Truth = Slope × Step + Curvature):

$$\mathcal{L}_k(\theta^\star + \Delta) - \mathcal{L}_k(\theta^\star) \;=\; \langle \nabla \mathcal{L}_k(\theta^\star), \Delta \rangle + \underbrace{\frac{1}{2W_k}\Delta^\top \hat{V}_k \Delta}_{\mathcal{Q}_k(\Delta)} .$$

Substituting this identity into the optimality condition and rearranging terms to isolate the quadratic form $\mathcal{Q}_k(\Delta)$ on the left-hand side:

$$\frac{1}{2W_k}\Delta^\top \hat{V}_k \Delta \;\le\; \alpha_k \left(\|\theta^\star\|_1 - \|\theta^\star + \Delta\|_1\right) - \langle \nabla \mathcal{L}_k(\theta^\star), \Delta\rangle \,. \tag{18}$$

We assume the good event where the regularization parameter satisfies $\alpha_k \ge 2\left\|\nabla \mathcal{L}_k(\theta^\star)\right\|_\infty$. Using Hölder's inequality on the inner product:

$$|\langle \nabla \mathcal{L}_k(\theta^\star), \Delta\rangle| \le \|\nabla \mathcal{L}_k(\theta^\star)\|_\infty \|\Delta\|_1 \le \frac{\alpha_k}{2}\|\Delta\|_1 \,.$$

Substituting this bound into 18:

$$\frac{1}{2W_k}\Delta^\top \hat{V}_k \Delta \;\le\; \alpha_k \left(\|\theta^\star\|_1 - \|\theta^\star + \Delta\|_1\right) + \frac{\alpha_k}{2}\|\Delta\|_1 \,.$$

We decompose the ℓ_1 norm of Δ into components on the support S and off the support S^c. Note that $\|\theta^\star\|_1 = \|\theta^\star_S\|_1$ and $\|\theta^\star + \Delta\|_1 = \|\theta^\star_S + \Delta_S\|_1 + \|\Delta_{S^c}\|_1$.

Using the reverse triangle inequality $\|\theta^\star_S\|_1 - \|\theta^\star_S + \Delta_S\|_1 \le \|\Delta_S\|_1$, the term in the parentheses becomes:

$$\|\theta^\star\|_1 - \|\theta^\star + \Delta\|_1 \le \|\Delta_S\|_1 - \|\Delta_{S^c}\|_1 \,.$$

Substituting this back:

$$0 \le \frac{1}{2W_k}\Delta^\top \hat{V}_k \Delta \;\le\; \alpha_k(\|\Delta_S\|_1 - \|\Delta_{S^c}\|_1) + \frac{\alpha_k}{2}(\|\Delta_S\|_1 + \|\Delta_{S^c}\|_1).$$

Ignoring the non-negative quadratic form for a moment gives us the cone constraint:

$$0 \le \frac{3\alpha_k}{2}\|\Delta_S\|_1 - \frac{\alpha_k}{2}\|\Delta_{S^c}\|_1 \implies \|\Delta_{S^c}\|_1 \le 3\|\Delta_S\|_1 \,.$$

Thus, the error vector Δ lies in the cone $C(S, 3)$.

Since $\Delta \in C(S, 3)$, we invoke Assumption 2 (Restricted Eigenvalue condition), which guarantees $\frac{1}{W_k}\Delta^\top \hat{V}_k \Delta \ge \kappa^2 \|\Delta\|_2^2$. Returning to the inequality from Step 3:

$$\frac{\kappa^2}{2}\|\Delta\|_2^2 \;\le\; \frac{1}{2W_k}\Delta^\top \hat{V}_k \Delta \;\le\; \frac{3\alpha_k}{2}\|\Delta_S\|_1 \,.$$

Using the norm inequality $\|\Delta_S\|_1 \le \sqrt{s^\star}\|\Delta\|_2$ [1]:

$$\frac{\kappa^2}{2}\|\Delta\|_2^2 \;\le\; \frac{3\alpha_k\sqrt{s^\star}}{2}\|\Delta\|_2 \,.$$

Dividing by $\|\Delta\|_2$ (assuming $\Delta \ne 0$) yields the L_2 error bound:

$$\|\Delta\|_2 \le \frac{3\sqrt{s^\star}}{\kappa^2}\alpha_k.$$

Finally, to get the L_1 bound required for the confidence radius, we use the cone property $\|\Delta\|_1 \le 4\|\Delta_S\|_1 \le 4\sqrt{s^\star}\|\Delta\|_2$:

$$\left\|\hat{\theta}_k - \theta^\star\right\|_1 = \|\Delta\|_1 \le 4\sqrt{s^\star}\left(\frac{3\sqrt{s^\star}}{\kappa^2}\alpha_k\right) = \frac{12 s^\star}{\kappa^2}\alpha_k.$$

This completes the proof.

[1] The norm inequality $\|\Delta_S\|_1 \le \sqrt{s^\star}\|\Delta\|_2$ follows directly from the Cauchy-Schwarz inequality. By viewing the L_1 norm on the support S as the inner product between the vector of absolute errors $(|\Delta_j|)_{j\in S}$ and the all-ones vector $\mathbf{1}_S$, we have $\|\Delta_S\|_1 = \sum_{j\in S}|\Delta_j|\cdot 1 \le (\sum_{j\in S}|\Delta_j|^2)^{1/2}(\sum_{j\in S}1^2)^{1/2} = \|\Delta_S\|_2\sqrt{|S|}$. Since $|S| \le s^\star$ and $\|\Delta_S\|_2 \le \|\Delta\|_2$, the bound holds.

C Proof of cumulative regret

We provide the detailed proof of Theorem 1.

Step 0: Good Events. Define $\mathcal{E}$ as the intersection of the Quantum Mean Estimation success event ($\mathcal{E}_{\mathrm{QME}}$) and the Restricted Eigenvalue success event ($\mathcal{E}_{\mathrm{RE}}$).

- By the union bound over epochs with $\delta_k = \frac{6\delta_{\mathrm{tot}}}{\pi^2 k^2}$, we have $\mathbb{P}(\mathcal{E}_{\mathrm{QME}}) \geq 1 - \delta_{\mathrm{tot}}$.
- By Assumption 2, $\mathbb{P}(\mathcal{E}_{\mathrm{RE}}) \geq 1 - \delta_{\mathrm{RE}}$.

Thus, with probability at least $1 - \delta_{\mathrm{tot}} - \delta_{\mathrm{RE}}$, all lemmas hold for all epochs.

Step 1: Validity of Confidence Sets. Conditioned on $\mathcal{E}$, Lemma 1 guarantees that the estimation error is bounded by the confidence radius β_{k-1} at every epoch k. Therefore, $\theta^\star \in \mathcal{C}_{k-1} = \{\theta : \left\|\theta - \hat{\theta}_{k-1}\right\|_1 \leq \beta_{k-1}\}$.

Step 2: Bounding the Gap via "Safe" Optimism. We must bound the sub-optimality of the played base action x_k. Note that x_k maximizes the UCB index $U_{k-1}(\cdot)$ over the *safe set* $\mathcal{A}' = (1-\gamma)\mathcal{A}$, not the full set $\mathcal{A}$.

Let $x^\star$ be the optimal action in the full set $\mathcal{A}$. Since $\mathcal{A}$ is a convex body containing the origin, the shrunken vector $x'_\star := (1-\gamma)x^\star$ lies in the safe set $\mathcal{A}'$.

1. **Optimism on Safe Set:** Since x_k is the maximizer over $\mathcal{A}'$, it must have a higher index than the shrunken optimal arm:

$$U_{k-1}(x_k) \geq U_{k-1}(x'_\star).$$

2. **Scaling Property:** The UCB index is homogenous (for $c \geq 0$).

$$U_{k-1}(cx) = cx^\top \hat{\theta}_{k-1} + \beta_{k-1}\left\|cx\right\|_\infty = cU_{k-1}(x).$$

   Therefore, $U_{k-1}(x'_\star) = (1-\gamma)U_{k-1}(x^\star)$.

3. **Standard Optimism:** Since $\theta^\star \in \mathcal{C}_{k-1}$, we have $U_{k-1}(x^\star) \geq (x^\star)^\top\theta^\star$. Combining these:

$$U_{k-1}(x_k) \geq (1-\gamma)(x^\star)^\top\theta^\star.$$

4. **Lower Bound on Played Action:** By definition of UCB and the confidence bound:

$$x_k^\top\theta^\star \geq x_k^\top\hat{\theta}_{k-1} - \beta_{k-1}\left\|x_k\right\|_\infty = U_{k-1}(x_k) - 2\beta_{k-1}\left\|x_k\right\|_\infty.$$

   Since $x_k \in \mathcal{A}'$, $\|x_k\|_\infty \leq 1-\gamma \leq 1$. Thus $x_k^\top\theta^\star \geq U_{k-1}(x_k) - 2\beta_{k-1}$.

Combining (3) and (4):

$$x_k^\top\theta^\star \geq (1-\gamma)(x^\star)^\top\theta^\star - 2\beta_{k-1}.$$

Rearranging for the gap:

$$\mathrm{Gap}_k = (x^\star)^\top\theta^\star - x_k^\top\theta^\star \leq 2\beta_{k-1} + \gamma(x^\star)^\top\theta^\star.$$

By choosing γ sufficiently small (e.g., $\gamma \approx 1/Q_{\mathrm{total}}$), the γ-term becomes negligible compared to the estimation error. In the $\widetilde{\mathcal{O}}$ analysis, we treat $\mathrm{Gap}_k \lesssim 2\beta_{k-1}$.

Step 3: The Quantum Cancellation. The regret for epoch k is $n_k \cdot \mathrm{Gap}_k$.

- **Error:** $\beta_{k-1} = \widetilde{\mathcal{O}}\left(\frac{s^\star}{\kappa^2}\sqrt{\frac{\log d}{W_{k-1}}}\right)$.
- **Cost:** We set $\varepsilon_k = 1/\sqrt{W_{k-1}}$. The quantum query complexity is $n_k = \widetilde{\mathcal{O}}(\frac{1}{\varepsilon_k}) = \widetilde{\mathcal{O}}(\sqrt{W_{k-1}})$.

Multiplying these terms cancels the W_{k-1} factor:

$$\mathrm{Regret}_k \approx n_k \beta_{k-1} = \widetilde{\mathcal{O}}\left(\sqrt{W_{k-1}} \cdot \frac{1}{\sqrt{W_{k-1}}}\right) = \widetilde{\mathcal{O}}\left(\frac{s^\star \sqrt{\log d}}{\kappa^2}\right).$$

Step 4: Summation over Epochs. The weights double every epoch ($W_k \approx 2W_{k-1}$), so W_k grows geometrically. The total query budget determines the number of epochs K:

$$Q_{\mathrm{total}} \approx \sum_{k=1}^{K} n_k \approx \sum_{k=1}^{K} \sqrt{2^k} \approx \sqrt{2^K}.$$

Thus, $K \approx \log_2(Q_{\mathrm{total}}^2) \approx \log(Q_{\mathrm{total}})$. Summing the constant regret per epoch over K epochs:

$$R(Q_{\mathrm{total}}) \leq \sum_{k=1}^{K} \mathrm{Regret}_k = \widetilde{\mathcal{O}}\left(\frac{s^\star \sqrt{\log d}}{\kappa^2} \log Q_{\mathrm{total}}\right).$$

This completes the proof.